DOCUMENT RESUME

ED 077 995 TM 002 807

AUTHOR	Flaugher, Ronald L.
TITLE	Some Points of Confusion in Discussing the Testing of Black Students.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RM-73-5
PUB DATE	Mar 73
NOTE	10p.; Paper prepared for a Symposium of the American Educational Research Association Annual Meeting (New Orleans, Louisiana, February 26, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Communication Problems; *Educational Testing; *Negro Students; *Psychometrics; Racial Discrimination; Speeches; Standardized Tests; *Test Bias; Test Interpretation; Test Validity

ABSTRACT

Four confusing issues that have delayed progress toward an awareness that testing is not a source of unfairness for minority students are discussed: (1) the assumptions underlying most of our psychometric manipulations are often not acknowledged or understood; (2) the extent of the objectivity of psychometrics is frequently exaggerated; (3) the meaning of certain terms, particularly "validity" (largely because it has both a technical and common usage), is quite confused; and (4) the understanding of just what function the tests are serving shifts from one function to another, unnoticed by those concerned. (KM)

ED 077995

TM 002 807

RM-73-5

RESEARCH MEMORANDUM

SOME POINTS OF CONFUSION IN DISCUSSING THE TESTING OF BLACK STUDENTS

Ronald L. Flaugher

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

This paper was presented at a symposium on "The Testing of Black Students," at the Annual Meeting of the American Educational Research Association, New Orleans, Louisiana, February 26, 1973.

Educational Testing Service
Princeton, New Jersey
March 1973

FILMED FROM BEST AVAILABLE COPY

SOME POINTS OF CONFUSION IN DISCUSSING THE TESTING OF BLACK STUDENTS[1]

Ronald L. Flaugher
Educational Testing Service

There are several confusing issues that have delayed the progress toward seeing that testing is not a source of unfairness for minority students. In my opinion, four such issues predominate as sources of this confusion: first, the particular assumptions which underlie most of our psychometric manipulations are often not acknowledged or understood; second, the extent of the objectivity of psychometrics is frequently exaggerated; third, confusion is rife over the meaning of certain terms--in particular, that of "validity"; and fourth, the issue I consider both most important and most difficult to handle is the shifting of the understanding of just what function the tests are serving. This shift from one function to the next frequently goes unnoticed during the course of a discussion, and confusion results when various participants in the dialogue attempt to deal with problems that accompany particular functions that they have assumed are primary, while others attempt to deal with other problems, stemming from other functions. By pointing them out, perhaps we can circumvent some of these sources of confusion. But let me discuss each of the four issues in turn.

First, I believe that there is some misunderstanding about the nature of the assumptions on which the psychometric model is based. The most usual circumstance is one in which some selection must take place, a selection based on a prediction of how well the student will perform on some criterion measure, such as the grade point average in college. In this setting, further, the

---

[1]Paper presented at a symposium on "The Testing of Black Students," at the 1973 Annual Meeting of the American Educational Research Association, New Orleans, La., February 26, 1973.

meritocratic principle usually applies, in that those who are predicted to do best, by whatever means the prediction is made, are the ones who are given top priority. There are other possible principles which could be used, but the meritocratic is by far the most frequently employed, although often without being acknowledged. Finally, there is a certain fixedness about the criterion in this psychometric technique, in the sense that once the criterion is decided upon, the psychometrics then have the job of doing the best possible job of predicting it; the model itself has no place in it for somehow evaluating the criterion once it has been accepted.

In any one of these instances, the potential for misunderstanding exists. If some discussants do not accept that some candidates are to be selected--and, consequently, that some candidates are to be rejected--or do not agree on the use of the meritocratic principle, or dispute the appropriateness of the criterion measure, then this should be made clear. When the real disagreement is over these basic assumptions, discussions about the "fairness" of the test content is futile.

A second major misunderstanding among discussants of the issue of testing minority students involves just how objective, in the last analysis, any psychometric selection system can be. It has now become obvious, as a result of the contributions of Thorndike (1971) and Darlington (1971), that there is never going to be a universally accepted, completely objective determination of the fairness of a test used as a selection device (Cole, 1972; Linn, 1973). No longer is it possible to resort to statistics for an impersonal completion of the selection decisions; rather, value judgments must be made explicit, and statistics can only be used as a means to implement those values, to put them into practice, once they have been established. So even the objective statistical

approaches must be preceded by a very subjective determination of what constitutes just and fair selection practices. A few years ago, we thought we had this model as the court of last appeal, but it is now clear that we were overlooking the existence of these alternative and conflicting interpretations of what constitutes fairness.

These developments are proving awkward because, for example, the Equal Employment Opportunities Commission's Guidelines on Employee Selection Procedures (EEOC, 1970) were established before this realization had occurred, and these established policies are in conflict with what we now understand to be the situation. It may be some considerable time before this particular point is no longer an obstacle to clear communication.

Another source of misunderstanding arises from the fact that many criticisms of existing testing practices are made by persons who do not use the same terminology as the test specialist. The psychometrician often has rigidly precise, and perhaps too narrow definitions, while the lay critic is operating from a "gut-level" knowledge that minority group members are not being treated fairly by society and that testing plays a role in this process. The difficulty arises when the psychometrician attempts a technical explanation of his understanding of the problem; the language and concepts he employs are often either not accessible to the layman or are used in different ways, and causes the layman to see the response as evasiveness on the psychometrician's part rather than a sincere attempt at communication. On the other hand, the psychometrician frequently sees the layman's rejection of his attempts as evidence that the layman "isn't really trying to understand." And hard feelings on both sides are the only result.

A good example of what I mean is the concept of "validity." Whenever a term has both a technical and a common usage--and "validity" is such a term--there is a potential for confusion. When laymen proclaim with absolute certainty that a test is "not valid," they may very possibly be using the term in a way that does not correspond at all to the technical use of the term; so when a psychometrically sound validation study is conducted, one which demonstrates that the tests are in fact technically "valid," this evidence is not seen by the layman as refuting the accusation. To a layman, a test is "not valid" if he knows, or knows of, someone who was turned away from an opportunity on the basis of test scores, but who somehow circumvented the barriers, went on, and succeeded. Any procedure that turns away someone who would have succeeded is invalid, in these terms. But this kind of "proof" of the invalidity of a test is quite compatible, of course, with a simultaneous demonstration of adequate predictive validity by the psychometric definition. Even the most precise of predictive measures necessarily have their share of cases for which incorrect predictions of failure are made, and this occurs regardless of ethnic group membership; for that matter, there are always errors of the opposite sort, falsely predicting success on the criterion. Such is the state of the art of academic prediction, and this may well be a primary source of misunderstanding between minorities, test makers, and admissions officials. Ironically, the difficulty arises from an exaggerated impression on the part of nontechnicians about just how effective testing could be, rather than in a belief that tests are valueless. In the lay terms, any deviation from flawless prediction is proof of invalidity, while the psychometrician has no such hopes for his methods.

The fourth and perhaps most serious source of confusion is the unnoticed shifting that occurs across several perceived functions for test information.

Depending upon which of these functions is being assumed, the same test data can be interpreted quite differently; the problem is that the functions are somewhat contradictory, they cannot in most cases be served simultaneously, and very different conclusions can be drawn about what is fair and unfair.

One of these functions I have already discussed in some detail, that of the prediction of some criterion for the purposes of selection and, perhaps, guidance; there are two other possible functions.

One is that of educational accountability. In this role, the tests are the measuring instruments which describe the outcome of a treatment, such as a year in school. They might be the means to determine, for example, whether or not a school has successfully taught the children to read. When tests are serving this function, they provide the objective evidence necessary to hold the schools accountable for the job they are doing, and as such, far from being a part of the problem, tests are an absolutely essential part of the solution. It is seldom the case that those who call for the elimination of testing altogether really mean that they are willing to allow the educational system to be released from any accountability at all, yet this would be one of the consequences.

But another function for tests exists, one which has no official status in the sense that prediction and accountability have, though it may be the most significant source of misunderstanding. Whether it is intended or not, in our society test information is frequently used as an <u>index of personal worth</u>. When this interpretation is put on test scores, then the same low mean score for a minority student is seen, not as an indication of potential difficulty in college, as in the prediction function, nor as an indication that the educational system has failed to do its job, as in the accountability function. Rather, the low score is seen as an attempted condemnation of minorities as a group by the

establishment, an attempt to certify that these groups are somehow of less worth. Even if this misunderstanding were the only one, it would still be sufficient to halt any cooperation between the two factions.

It is important to study this matter closely, because it may well be the source of most of the emotionality that has slowed cooperation and progress in the past. Certainly, predictive validity coefficients are not that controversial, once the ground rules are agreed upon, although as I have indicated, that might be a problem in itself. Similarly, the disagreements do not seem to be generated within the educational accountability framework, at least not with those who are being tested; any arguments are usually with the educators, and the unfairness question is about whether the test really measures what they were attempting to teach. Some emotionality occurs, but it is usually not on the part of minority spokesmen.

The strongest disagreements, I believe, occur when this unofficial "personal worth" function of tests comes into play; that is, when one or both of the parties to the discussion believes that the tests provide evidence of worth rather than predictive or accountability information. No one can, or should, accept such an interpretation quietly. Unfortunately, however, when critics demand that the source of such unacceptable interpretations be eliminated, the quite legitimate functions of prediction and accountability are lost from consideration. Further, those discussants who continue to think only in terms of those two legitimate functions are lost, too, for they fail to understand what the objections are all about. To them, objecting to a statistical conclusion about predictive power seems inappropriate, and denouncing the accountability function seems counterproductive, so, ignorant of the true nature of the objection, they are left quite puzzled.

A good example of this confusion of function, I think, lies in the distinction between IQ tests and aptitude and achievement examinations. In a sense there is, in fact, a certain "personal worth" interpretation that is put on the results of an IQ test, and this is somewhat encouraged by the more general, vague nature of the purposes for which it is given. This more diffuse nature of the purpose does, in fact, lend itself to a personal worth interpretation. Contrast this with the college aptitude examinations, which have a distinct function to fill and which can constantly be checked and verified to see if a good job is being done. Or consider the achievement test, the content of which has to be agreed upon by the subject matter specialist, and can be discarded or revised in response to qualified opinion. The "IQ," however, has much more of a final, unquestioned and innate sort of sound to it, and this one may well be the source of the trouble. Perhaps much of the misunderstanding can be eliminated if we vow to make clear the distinctions among these types of tests, and perhaps we will be able to keep the underlying functions they serve distinct as well.

Then, when we call for modifications of testing practices, we can do so in a way that doesn't eliminate those positive functions that testing does serve, indeed functions which I believe minority students can ill afford to do without.

So my suggestion for a good beginning to any discussion concerning the testing of minority students is to check these points:

First, is it agreed that the problem is to be one of selection of a few students from among many? Is it to be an instance of the meritocratic principle? And what about the criterion--can we agree on what it is that we want to predict?

Secondly, are we aware that no matter which system we use, we are imposing a particular system of values, whether openly expressed or not, and that the psychometric manipulations are not somehow an escape from such value systems?

Third, are we clear about the terminology we are using? In particular, is it clear when we are using the technical meaning of "valid" and when the layman's sense of the word?

Finally, are we aware of the various functions that tests serve in our society, and are we perhaps confusing academic prediction with educational evaluation, and both of these with an index of personal worth?

If we settle these issues, I think we will be well on our way to achieving that fairness which we seek.

References

Cole, N. S. Bias in selection. ACT Research Report No. 51. Iowa City, Iowa: American College Testing Program, 1972.

Darlington, R. D. Another look at "culture fairness." Journal of Educational Measurement, 1971, 8, 71-82.

Equal Employment Opportunity Commission, Part 1607--Guidelines on employee selection procedures. Code of Federal Regulations, Title 29, Chapter XIV, 1970.

Linn, R. L. Fair test use in selection. Review of Educational Research, 1973 (in press).

Thorndike, R. L. Concepts of culture fairness. Journal of Educational Measurement, 1971, 8, 63-70.